Chapter 3

A COMPLEX HISTORY OF PHOTOTROPHY REVEALED BY NOVEL CHLOROFLEXI LINEAGES

Ward, Lewis M., James Hemp, Patrick M. Shih, Shawn E. McGlynn, and Woodward W. Fischer. In preparation.

Abstract:

The antiquity of photosynthesis is a point of substantial contention, with competing hypotheses for either its presence in the last universal common ancestor and gene loss in most lineages, or a much later origin followed by extensive horizontal gene transfer into the extant phototrophic clades. Selecting between these alternatives requires constraining whether photosynthesis-associated genes have been inherited vertically from a common photosynthetic ancestor, or whether they have been horizontally transferred. While the phylogenetic relationships between the bacterial phyla are unclear, making the order of acquisition of photosynthesis uncertain, much greater resolution is retained within phyla, where more recent cases of vertical inheritance or horizontal gene transfer can be tested. Here, we report several new draft genome sequences from within the Chloroflexi phylum that form a metabolically diverse, monophyletic clade sister to the Anaerolineae class that we term *Candidatus* Thermofonseaa. This class includes two independently phototrophic lineages. Comparison of organismal (based on conserved ribosomal proteins) and phototrophic gene (based on reaction center and bacteriochlorophyll synthase) trees demonstrate that these lineages acquired phototrophy via horizontal gene transfer from the phototrophic Chloroflexia class. The extent of horizontal gene transfer of phototrophy, as

well as other metabolic pathways, within the Chloroflexi is indicative of a broader role for horizontal transfer in the distribution of metabolic traits throughout the bacteria, challenging assumptions about metabolism in the last universal common ancestor and subsequent innovation and diversification of metabolic pathways.

Introduction:

Anoxygenic photosynthesis is among the oldest and most important metabolic inventions in the history of life on Earth. Anoxygenic photosynthesis was a critical driver of primary productivity on the early Earth, and it gave rise to oxygenic photosynthesis, which revolutionized biology and geochemistry and ultimately fueled the rise of complex life. Multiple hypotheses exist for the origin and subsequent evolution of anoxygenic photosynthesis, but little is known with certainty. To date, chlorin-based phototrophy has been identified in seven bacterial phyla: the Cyanobacteria, Chlorobi, Chloroflexi, Acidobacteria, Heliobacteria, Gemmatimonadetes, and Proteobacteria (Figure 1). Of these, only one—the Cyanobacteria—possesses two photosystems coupled to perform oxygenic photosynthesis. The others perform anoxygenic phototrophy and possess only a single reaction center, either Type 1 (Chlorobi, Heliobacteria, and Acidobacteria) or Type 2 (Proteobacteria, Gemmatimonadetes, and Chloroflexi). While it has been suggested that photosynthesis was present in the last common ancestor of the bacteria (Woese et al. 1985, Woese 1987), followed by extensive loss in most lineages, this idea remains controversial. An alternative scenario involves a later origin of photosynthesis, followed by multiple instances of horizontal gene transfer (HGT) resulting in the modern distribution of photosynthesis (e.g. Raymond et al. 2002).

A test of these alternatives is to compare the organismal phylogenies of phototrophic bacteria to the gene phylogenies of photosynthesis genes—concordance of the trees would indicate shared ancestry, while discrepancies between them would indicate a history of horizontal gene transfer (Doolittle 1986). While the structure of the bacterial tree of life is debated (e.g. Woese 1987, Williams et al. 2013, McInerney et al. 2014), intra-phylum organismal relationships are generally robust despite uncertainty in relationships between phyla (Pace 2009). As a result, the history of photosynthesis within a phylum is more straightforward to assess than it is for the bacteria as a whole. If a major role for horizontal gene transfer can be demonstrated within a particular phylum, the HGT-driven phototrophy hypothesis will be strengthened, whereas a concordance of organismal and gene trees would be more consistent with an ancient origin and vertical inheritance of the metabolism. While tests of this kind have been made previously in the Proteobacteria, suggesting intra-phylum horizontal gene transfer (Igarashi et al. 2001, Nagashima & Nagashima 2013), this has not previously been possible in other phototrophic phyla due to the limited diversity of phototrophic members within each. However, new opportunities for assessing the history of photosynthesis within the Chloroflexi phylum has recently been made possible by the discovery of novel lineages of phototrophs within this phylum.

The Chloroflexi (i.e. Green Nonsulfur Bacteria) are a phylum of primarily gliding, filamentous bacteria possessing a wide diversity of metabolisms and ecological roles, but are best known as photoheterotrophs (Overmann 2008). Chloroflexi have been shown to be diverse and abundant in a range of environments (e.g. marine sediments and groundwater, Inagaki et al. 2003, Hug et al. 2013). Despite their environmental richness revealed by

culture-independent surveys, most described Chloroflexi belong to a few subclades isolated from hot springs (Yamada and Sekiguchi 2009), including the anoxygenic phototrophic *Chloroflexus* (Pierson and Castenholz 1974, Hanada et al. 1995). Phylogenetic analysis of the phototrophy genes of Chloroflexi has suggested that anoxygenic photosynthesis in this group predates the evolution of oxygenic photosynthesis in Cyanobacteria (Xiong et al. 2000), implying that this group is ancient and therefore a good candidate for investigating questions of early Earth history. Recent culture- and sequence- based efforts have expanded the known taxonomic and metabolic diversity of the Chloroflexi phylum (e.g the Ardenticatenia class, capable of nitrate- and iron oxide- reduction, Hemp et al. 2015b, Kawaichi et al. 2015). Newly discovered Chloroflexi vary tremendously in their morphology, metabolism, and other traits (Table 1), but are recovered as a monophyletic clade in phylogenetic trees (Figure 2) and have sufficient sequence similarity to be classified as a single phylum (Hanada 2014).

Here, we report draft genomes of three lineages of phototrophic Chloroflexi, including two outside of the classically phototrophic Chloroflexia class, and demonstrate the incongruence of organismal and phototrophic gene trees, suggesting a history of horizontal gene transfer of photosynthesis within this phylum. Moreover, multiple genome bins were recovered for nonphototrophic relatives of two of these novel phototrophic Chloroflexi, forming a new clade sister to the Anaerolineae class of Chloroflexi. This new class-level clade was recovered as several genome bins from Japanese hot spring metagenomes, including members with genes for aerobic and anaerobic respiration as well as both chlorin- and rhodopsin-based phototrophy. The discovery of diverse members of an

entire new class of Chloroflexi from just two Japanese hot springs suggests a large amount of unknown diversity in this phylum that may be recovered by further culture- and sequencing-based efforts.

Methods:

Metagenomic sample collection:

Four metagenomic datasets were recovered from two hot springs in Japan, Jinata Onsen and Nakabusa Onsen. Genome bins labeled JP1 or JP3 were derived from Jinata Onsen, while CP1 and CP2 were derived from Nakabusa Onsen.

Jinata genome bins were assembled from two metagenomes from Jinata Onsen, on Shikinejima Island, Tokyo Prefecture. The geochemistry and microbial diversity of this spring is described in detail elsewhere (Ward et al. 2017a). Jinata Onsen is located at 34.326111N, 139.21E on the island of Shikinejima, Tokyo Prefecture, Japan. Shikinejima is part of the Izu Islands, a chain of volcanic islands that formed in the past 2-3 million years along the northern edge of the Izu-Bonin-Mariana Arc (Kaneoka et al. 1970). The source water of Jinata Onsen emerges anoxic, iron-rich, and gently bubbling from the spring source. Temperatures at the source are ~62°C. This spring water flows into a series of pools that mix progressively more with seawater during high tide, creating a range of geochemical conditions over short spatial and temporal scales as hot, iron-rich, oxygen-poor spring water mixes with cold, sulfate- and oxygen-rich seawater. The metagenomes from which JP1 bins were recovered was sequenced from an iron-oxide rich pool near the spring source (Pool 1), while JP3 genomes were recovered from a Cyanobacteria-rich

microbial mat in Pool 3, the most downstream section of the hot spring before it flows into the open ocean. Dissolved oxygen (DO), pH, and temperature measurements were performed *in situ* using an Exetech DO700 8-in-1 Portable Dissolved Oxygen Meter. Iron concentrations were measured using the ferrozine assay (Stookey 1970) following acidification with 40 mM sulfamic acid to inhibit iron oxidation by O_2 or oxidized nitrogen species (Klueglein and Kappler 2013). At the time of sampling, Pool 1 59°C, pH 5.8, and contained 1.8 mg/L dissolved oxygen and 265 μM Fe^{2+}; Pool 3 at Jinata Onsen was 46°C, pH 6.7, and contained 5.6 m/L dissolved oxygen and 100 μM Fe^{2+}.

Nakabusa genome bins were assembled from two metagenome from hot spring microbial mats from Nakabusa Onsen, located at 36.392429N, 137.748038E in the Japanese Alps near Azumino, Nagano Prefecture. Geochemical and microbiological characterization of Nakabusa Onsen is described in detail elsewhere (Kubo et al. 2011, Ward et al. 2017b). Nakabusa Onsen is a sulfidic, moderately alkaline hot spring with source waters near 70°C. The samples from which the metagenomes were sequenced were of cone-forming microbial mats at two points along the outflow from the hot spring source; Cone Pool 1 (the source of CP1 genomes) was a Chloroflexi-dominated mat near the hot spring source, which at the time of sampling was 48°C and pH 8.1, while Cone Pool 2 (the source of the CP2 genomes) was collected from a cone-forming Cyanobacteria-rich microbial mat several meters downstream, which at the time of sampling was 32°C and pH 8.3.

Samples of microbial mats were collected using sterile forceps and spatulas (~0.25 cm^3 of material). Cells were lysed and DNA preserved in the field using Zymo Terralyzer

BashingBead Matrix and Xpedition Lysis Buffer. Cells were disrupted immediately by attaching tubes to the blade of a cordless reciprocating saw and operating for 1 minute.

Metagenomic sequencing and analysis:

Following return to the lab, DNA was extracted and purified with a Zymo Soil/Fecal DNA extraction kit (Zymo Research, Irvine, CA). DNA was quantified with a Qubit 3.0 fluorimeter (Life Technologies, Carlsbad, CA) according to manufacturer's instructions following DNA extraction. Purified DNA was submitted to SeqMatic LLC (Fremont, CA) for library preparation and sequencing via Illumina HiSeq technology. Raw sequences were assembeled with MegaHit v. 1.02 (Li et al. 2016) and genome bins constructed using MetaWatt version 3.5.2 (Strous et al. 2012). Genomes were manually screened for genes of interest and uploaded to RAST (Aziz et al. 2008) for overall characterization. Genome bins were assessed for completeness and contamination using CheckM (Parks et al. 2014). Genes of interest (e.g. coding for ribosomal, photosynthesis, and electron transport proteins) were screened against outlier (e.g. likely contaminant) contigs as determined by CheckM using tetranucleotide, GC, and coding density content.

Phylogenetics

Sequences of ribosomal and phototrophy proteins were identified locally with BLAST+ (Camacho et al. 2008), aligned with MUSCLE (Edgar 2004), and alignments manually curated in Jalview (Waterhouse et al. 2009). Phylogenetic trees were calculated using RAxML (Stamatakis 2014) on the Cipres science gateway (Miller et al. 2010). Trees were visualized with Seaview (Gouy et al. 2010) and the Interactive Tree of Life (Letunic and Bork 2016).

Results and discussion:

Our sequencing efforts, including both hot spring metagenomes and sequencing of cultured isolates, have resulted in draft genomes of three new reaction center-containing phototrophic Chloroflexi lineages and eight genome bins which do not encode reaction centers but are associated with a new class-level clade sister to the Anaerolineae (Table 1, Figure 2). *Kouleothrix aurantiaca* represents a so-far monospecific genus within the class Chloroflexia, while JP3_7, CP2_42A, and the other genome bins reported here form a new clade sister to the Anaerolineae. Genome statistics for these bins are reported in Table 1, with summaries of the metabolic proteins encoded by these genomes in Table 2.

Organismal phylogenies of the Chloroflexi phylum, including the novel phototrophs and other genome bins described here, were constructed using the RpoB protein sequence (Figure 2). This protein is a core information processing protein, and is always found as a single copy that has been vertically inherited (Hansmann and Martin 2000), so the RpoB phylogeny should reflect the organismal phylogeny. For genomes in which a 16S gene was recovered, the 16S phylogeny matched the topology of the RpoB tree. In these phylogenies, *Kouleothrix aurantiaca* branched within the Chloroflexia class, basally to the *Roseiflexus* after their divergence from *Chloroflexus*. CP2_42A and JP3_7 were recovered as separate lineages, forming a clade sister to the Anaerolineae along with the other genome bins reported here.

Reaction center protein trees (Figure 3) show *Kouleothrix* in the same position relative to other Chloroflexia as in organismal trees, basal to the *Roseiflexus*, but place

CP2_42A and JP3_7 very differently—with CP2_42A as branching between *Kouleothrix* and *Roseiflexus*, and JP3_7 branching sister to the *Roseiflexus+CP2_42A+Kouleothrix* clade.

Candidatus Thermofonseae, a metabolically diverse novel class of Chloroflexi sister to Anaerolineae

The metagenome bins reported here, together with the "Anaerolineae-like" phototroph recovered from a Yellowstone National Park metagenome (Klatt et al. 2011), form a monophyletic clade sister to the Anaerolineae class in phylogenetic trees based on conserved organismal marker proteins such as RpoB (Figure 2). The members of this class appear to encode diverse heterotrophic metabolic traits, including photoheterotrophy and diverse pathways for both aerobic and anaerobic metabolism.

We propose for this clade the name *Candidatus* Thermofonseae, from the Latin for hot spring, with official classification pending isolation and characterization of at least one member. Genome bins falling within this clade were recovered from all four of our metagenomic datasets from both hot springs. The Thermofonseae genome bins reported here were up to ~96% complete as determined by single copy marker genes, and recovered diverse metabolic capabilities as described above, distinguishing them from their closest relatives, the metabolically-limited Anaerolineae. Of the ten genomes reported here, two include phototrophic reaction centers, two include rhodopsins, two possess partial denitrification pathways, and six contain genes for aerobic respiration (Supplemental Table 2). The Anaerolineae, in contrast, are typically described as obligate anaerobes (e.g. Yamada and Sekiguchi 2009), though genes for aerobic respiration have been recovered in

multiple Anaerolineae genomes (e.g. Pace et al. 2015, Ward et al. 2015a, Hemp et al. 2015c). Phylogenetic analysis of electron transport and respiration genes in the Thermofonseae and Anaerolineae reveal that metabolic protein trees are incongruent with organismal trees, suggesting independent acquisitions of respiration in these two clades (Supplemental Figures). The Thermofonseae utilize a bc complex for respiration, while the Anaerolineae commonly use an Alternative Complex III. Furthermore, the Heme Copper Oxidoreductases (HCOs) in these organisms are not closely related. Thermofonseae uses an A family HCO closely related to those of Cyanobacteria, while those in Anaerolineae are closely related to Caldilineae. This suggests that stem lineages of these classes diverged prior to the acquisition of aerobic respiration, followed by diversification after receiving this metabolism through horizontal gene transfer, or alternatively loss and replacement in one lineage.

At least three members of the Thermofonseae (CP2_42A, JP3_13, and CP2_2F) contain rhodopsin genes. Those of JP3_13 and CP2_2F are related to the "Actinorhodopsins" found in *Roseiflexus sp*. RS-1 which is thought to be functional as a light-driven proton pump (Sharma et al. 2008). The rhodopsins found in JP3_13 and CP2_2F have highly similar sequences; given the relatedness of these strains and their rhodopsins, these genes were likely inherited from the last common ancestor of these strains following acquisition via horizontal gene transfer. The rhodopsin encoded in the CP2_42A genome is most closely related to xanthorhodopsin, a proton-pumping rhodopsin shown to use light-harvesting antenna carotenoids (Balashov et al. 2005). Despite the presence of rhodopsins in diverse members of the Chloroflexi including *Roseiflexus*,

Ktedonobacter racemifer, and *Bellilinea caldifistulae* (members of the Chloroflexia, Ktedonobacteraceae, and Anaerolineae classes of the Chloroflexi, respectively), the rhodopsins in each of these Chloroflexi lineages are not closely related, and likely reflect independent acquisitions via horizontal gene transfer from other phyla and not a shared history of rhodopsins in the Chloroflexi phylum. Genes for the copper-containing nitrite reductase NirK are present in bins JP1_20 and CP2_20G, suggesting the potential for at least partial denitrification in these strains.

Anoxygenic phototrophs

Two members of *Ca*. Thermofonseae reported here, CP2_42A and JP3_7, contained genes indicate of anoxygenic phototrophy. The CP2_42A genome bin branches basally to the Anaerolineae class based on organismal trees built based on RpoB and concatenated ribosomal proteins. CP2_42A encodes genes for anoxygenic phototrophy; a type II reaction center (including a fused pufLM and a pufC), bacteriochlorophyll biosynthesis, a cytochrome bc complex, and Alternative Complex III.

JP3_7 is most closely related to the "Anaerolineae-like" phototrophic Chloroflexi assembled from a metagenome from Yellowstone National Park (Klatt et al. 2011), though it is genetically distinct to the species and possibly the genus level (~78% average nucleotide identity). JP3_7 encodes genes for anoxygenic phototrophy; a type II reaction center (including unfused pufL and pufM, and a pufC), bacteriochlorophyll synthesis, a cytochrome bc complex, but no Alternative Complex III. JP3_7 contains unfused PufL and PufM genes, similar to that in *Chloroflexus*.

Although CP2_42A and JP3_7 contain most genes involved in bacteriochlorophyll synthesis, including bchX,Y,Z,P, F,G,I,D, and a bchH homolog, both are missing bchL,N,B,E, or M. Neither CP2_42A nor JP3_7 bin recovered genes for the Calvin cycle or the 3-hydroxyprionate bicycle and so these strains may grow primarily as photoheterotrophs.

While the draft genomes reported here are largely too fragmented to recover informational genes on the same contigs as phototrophy related genes, the rpoB and bchP genes of JP3_7 are collocated on contig 8001, providing strong support for the inference of phototrophy in this lineage from the genome bins produced here.

Horizontal gene transfer of phototrophy within the Chloroflexi:

The position of *Kouleothrix* in both organismal and gene trees is consistent with a vertical inheritance of phototrophy from the last common ancestor of the *Roseiflexus+Chloroflexus* clade after its divergence from the nonphototrophic *Herpetosiphon*. However, the other two phototrophic Chloroflexi reported here reveal a more complicated history. These two strains (CP2_42A and JP3_7) branch within *Ca*. Thermofonseae in organismal trees based on conserved vertically inherited proteins like RpoB, quite distant from the other phototrophic Chloroflexi (Figure 2). However, photosynthesis-related gene trees place these strains within the other phototrophic Chloroflexi. CP2_42A branches within the Chloroflexia, basal to the clade of *Roseiflexus* and *Kouleothrix*. JP3_7, however, branches more deeply, sister to the *Roseiflexus+Kouleothrix+CP2_42A* clade (Figure 3). Furthermore, *Kouleothrix* and CP2_42A have fused pufL and pufM genes, a feature which appears in reaction centers of

Roseiflexus (Youvan et al. 1984, Yamada et al. 2005) and so appears to be a synapomorphy of this lineage of phototrophs, supporting their inclusion at this point in the phototrophy tree to the exclusion of JP3_7 which has unfused pufL and pufM genes.

The incongruence between organismal and gene trees for the novel phototrophic Chloroflexi described here suggests that photosynthesis genes were not vertically inherited from the last common ancestor of the phototrophic Chloroflexi. Instead, the differing branching order of JP3_7 and CP2_42A between RpoB and PufLM trees, along with the presence of a conserved gene fusion within the *Roseiflexus+Kouleothrix+CP2_42A* clade, strongly suggests that horizontal gene transfer has played a role in the current distribution of phototrophy in the Chloroflexi phylum.

In light of these data, the most parsimonious scenario for the evolution of phototrophy within the Chloroflexi requires at least two instances of horizontal gene transfer to have occurred (Figure 4). This scenario involves the acquisition of an unfused Type 2 reaction center in an ancestor of the Chloroflexia after the divergence of *Herpetosiphon*, horizontal gene transfer of this unfused ancestral form from the branch leading to *Roseiflexus* into the JP3_7 lineage, followed by a single pufL+pufM fusion event in an ancestor of *Roseiflexus* and *Kouleothrix*, and a second horizontal gene transfer event of the now fused protein into an ancestor of CP2_42A from the *Roseiflexus* lineage after its divergence from *Kouleothrix*.

Presence of other photosynthesis-related genes

The genome bins for CP2_42A and JP3_7 recover most, but not all, of the bacteriochlorophyll synthesis pathway expected for phototrophic Chloroflexi. These

genomes contain bchX,Y,Z,P,F,G,I,D, and a bchH-like gene, but not bchL,N,B, M, or E. While this may be a result of the incomplete nature of these genomes, the same bacteriochlorophyll synthesis gene complement has been described in the "Anaerolineae-like" phototroph genome bin recovered from a Yellowstone National Park metagenome (Klatt et al. 2011). Some or all of these genes may in fact actually be absent from these genomes, possibly functionally replaced by promiscuous homologs (e.g. bchLNB are homologous to bchXYZ, and chimeras of other homologs of these genes have been demonstrated to be functionally exchangeable, e.g. Wätzlich et al. 2009, Cheng et al. 2005). Isolation and biochemical characterization of the bacteriochlorophyll synthesis pathway in these organisms will be necessary to resolve this possibility. Estimates of the probability of missing the same set of genes from multiple genomes of relatively high (>50%) completeness are incredibly low (<<1%), strengthening interpretations of the absence of these genes from *Ca*. Thermofonseae genomes (Supplemental Information).

It is interesting to note that the vast majority of sequenced phototrophic Chloroflexi utilize Alternative Complex III (Yanyushin et al. 2005) for energy conservation during phototrophy, even to the extent of CP2_42A appearing to have acquired ACIII along with other phototrophy genes (Supplemental Information). However, ACIII was not recovered in the draft genomes for *K. aurantiaca* or JP3_7. This suggests that the use of ACIII for phototrophy may not by a synapomorphy of the phototrophic Chloroflexi, though this will require closure of these genomes and confirmation that ACIII is truly absent and not simply missing from the draft genome. The presence of auracyanin, the electron acceptor of ACIII (Majumder et al. 2013), in JP3_7 is consistent with the ancestral presence of ACIII in this

lineage and either recent loss or failure to recover the gene in the genome bin. Meanwhile, all of the aerobic members of the Thermofonseae encode a bc complex, consistent with other aerobic, nonphototrophic Chloroflexi clades such as Caldilineae and Ardenticatenia.

The history of carbon fixation in the Chloroflexi is complicated, and is not only not congruent with the organismal phylogeny but is not congruent with phototrophy gene phylogenies, suggesting that the light and dark reaction pathways for photosynthesis in the Chloroflexi have undergone independent histories of horizontal gene transfer. While the phototrophic Chloroflexi are well known to possess the 3-hydroxypropionate bicycle for carbon fixation (e.g. Berg 2011), this pathway is absent in the genomes reported here, as well as *Oscillochloris* and *Chlorothrix*. Instead, *Kouleothrix, Oscillochloris,* and *Chlorothrix* possess the Calvin Cycle, while CP2_42A and JP3_7 do not encode and carbon fixation pathways. Comparison of the phylogenies of these organisms and their phototrophy genes does not reveal a clear, consistent history for carbon fixation, with scenarios involving the first phototrophic Chloroflexi possessing the 3-HP bicycle, the Calvin cycle, or no carbon fixation at all being similarly parsimonious.

Conclusions:

The increased diversity of Chloroflexi phototrophs and history of HGT described here are consistent with other metabolic characters in this phylum. Our group has previously sequenced diverse representatives of the Chloroflexi, filling in gaps in the tree (Ward et al. 2015ab, Hemp et al. 2015abc, Pace et al. 2015) in order to better characterize the diversity and distribution of high potential metabolism within this phylum. Our previous efforts to expand the sequenced diversity of the Chloroflexi have revealed a high

degree of previously unrecognized metabolic diversity in this phylum, including high-potential metabolic pathways for aerobic and anaerobic respiration (Ward et al.2015ab, Hemp et al. 2015abc, Pace et al. 2015). The distribution and phylogenies of genes associated with these pathways is diverse and consistent with a history of horizontal gene transfer. Together, these data are consistent with a high degree of metabolic diversity and abundant horizontal gene transfer within the Chloroflexi phylum.

A history of horizontal gene transfer of phototrophy within the Chloroflexi is consistent with that of the Proteobacteria, which records extensive intra-phylum HGT (Igarashi et al. 2001, Nagashima & Nagashima 2013). A single clear case of inter-phylum HGT is also recorded in the presence of a Proteobacteria-derived RC2 in a member of the Gemmatimonadetes (Zeng et al. 2014).

An important question is the relative timing of acquisition of phototrophy in different lineages. If anoxygenic photosynthesis is an ancient metabolism, predating the introduction of molecular oxygen into biology as a result of oxygenic photosynthesis (e.g. Xiong et al. 2000), then at least some lineages should have acquired anoxygenic photosynthesis before the Great Oxygenation Event ~2.3 Gya, and therefore before the ability to respire oxygen. However, for many phototrophic groups, it appears that the acquisition of phototrophy postdated the acquisition of aerobic respiration (Fischer et al. 2016). As a result, the taxonomic affinity of the oldest anoxygenic phototrophs remains unclear. It is possible that photosynthesis originated in a member of a characterized phototrophic clade (e.g. the suggestion that phototrophy may have originated in the Chloroflexi, Oyaizu et al. 1987, or the Cyanobacteria, Mulkidjainian et al. 2006), within a

so-far undiscovered but still extant group, or may in fact have gone extinct. This can best be resolved by continued discovery of new phototrophic groups—an increasingly frequent phenomenon as environmental sequencing efforts continue and improve.

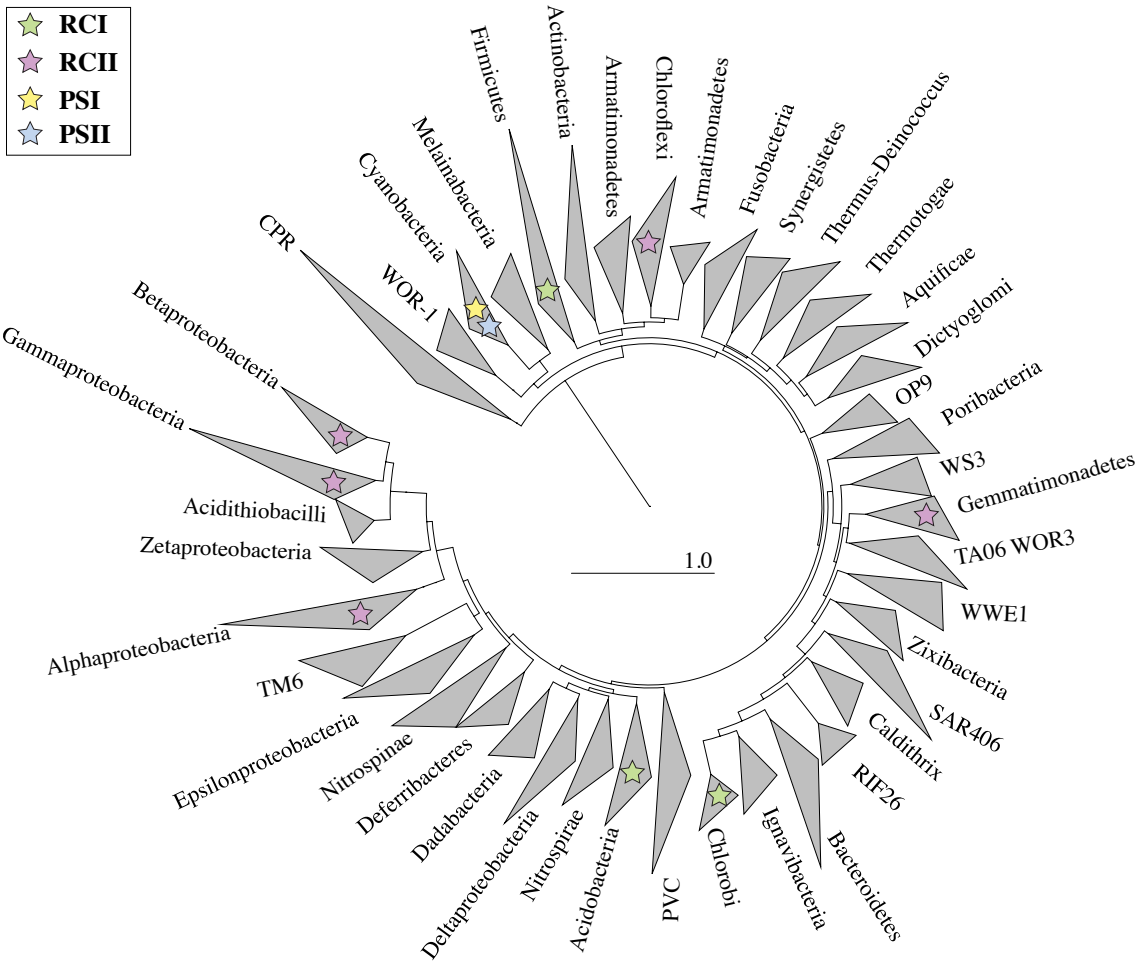


Figure 1: Phylogeny of bacteria based on concatenated ribosomal proteins, with clades that include phototrophs noted with symbols representing their reaction center. Modified from Hug et al. 2016 and visualized using the interactive tree of life (Letunic and Bork 2016).

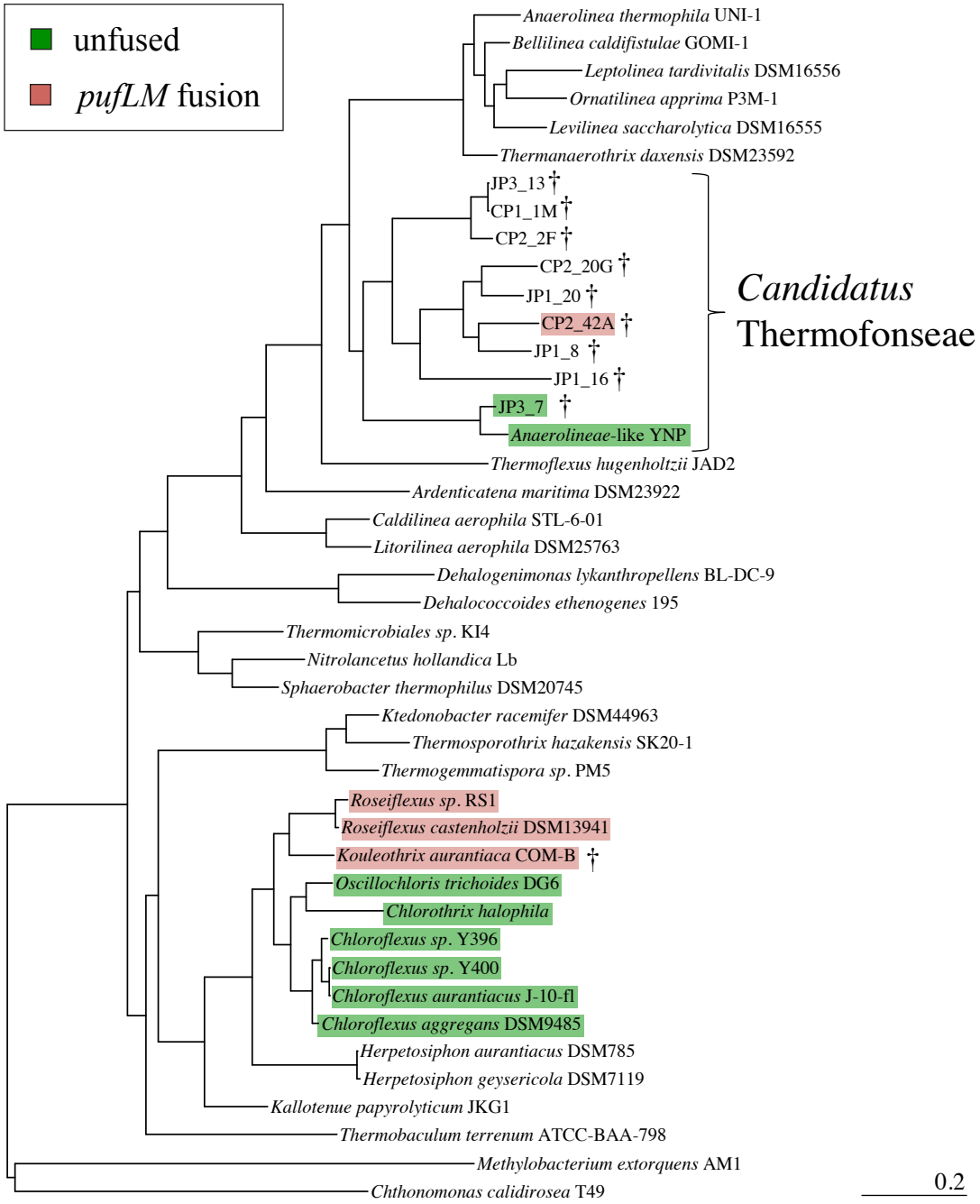


Figure 2: Phylogeny of Chloroflexi based on the RpoB protein sequence, with our newly sequenced strains indicated with daggers, phototrophic strains highlighted (red for fused pufLM, green for unfused), and *Candidatus* Thermofonseae noted.

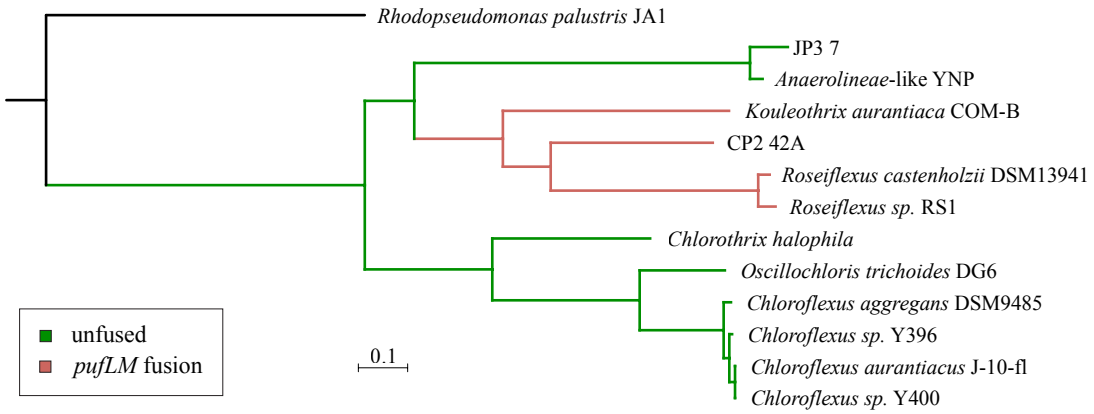


Figure 3: Phylogeny of Type 2 phototrophic reaction center proteins (concatenated PufL and PufM); lineages with fused PufLM proteins are highlighted in red, while lineages with unfused reaction centers are in green.

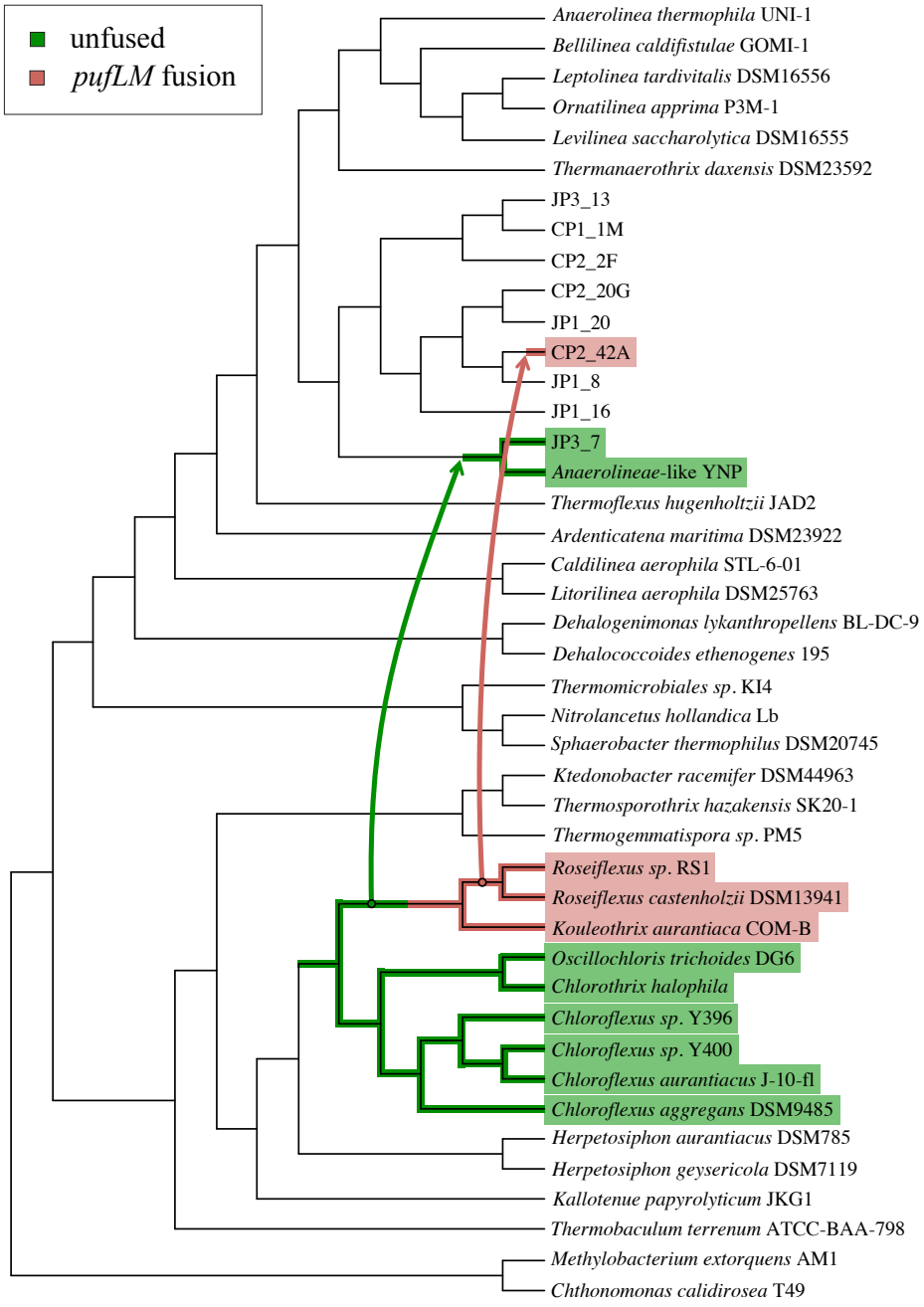


Figure 4: Cartoon of evolutionary scenario of phototrophy in Chloroflexi. Nonphototrophic lineages are in black, lineages with fused PufLM proteins are highlighted in red, and lineages with unfused reaction centers are in green. Arrows represent horizontal gene transfer of phototrophy genes.

	Chloroflexia	Thermomicrobia	Anaerolineae	Caldilineae	Ardenticatenia	Ktedenobacteria	Thermoflexia	Dehalococcoidetes	*Ca.* Thermofonseae
Phototrophy	+ (chlorin-based)	-	-	-	-	-	-	-	Some (chlorin- or rhodopsin-based)
Aerobic respiration	+	+	Genes present	Genes present	+	+	Microaerophilic	Genes rarely present	Genes present
Complex III	bc, ACIII, or both	ACIII or both	ACIII, bc, or neither	bc	bc	bc	bc	Neither	Bc, also ACIII with RCII
Morphology	Filamentous	Rod	Filamentous or rods	Filamentous	Filamentous	Filamentous (branched)	Filamentous	Coccoidal, discs	Unknown
Motility	Gliding	Flagellar	Flagellar or none	-	-	-	-	-	Unknown—no flagellar genes
Other metabolic traits		Nitrite oxidation			Iron and nitrogen respiration			Dehalogenation	Nitrogen respiration
Temperature range	10-67	43-80	20-73	37-65	30-75	17-74	67.5-75	15-35	32-59
% GC	48-62	56-63	48-58	59-65	51.5	54-60	69	49-54	46-63

Table 1: Characteristics of the classes of Chloroflexi, including *Ca*. Thermofonseae described here. The Chloroflexi phylum contains eight recognized classes, with genome sequences available for at least one member of each: Chloroflexia (Garrity & Holt, 2001; Gupta et al., 2013), Thermomicrobia (Hugenholtz&Stackebrandt, 2004, Sorokin et al.

2012), Ktedonobacteria (Cavaletti et al., 2006; Yabe et al., 2010; Chang et al. 2011), Dehalococcoidia (Löffler et al., 2013; Moe et al., 2009), Ardenticatenia (Kawaichi et al., 2013), Anaerolineae and Caldilineae (Yamada et al. 2006), and Thermoflexia (Dodsworth et al. 2014). Table data from this study, Dodsworth et al. 2014, and our previously described Chloroflexi genomes (Ward et al. 2015ab, Hemp et al. 2015abc, Pace et al. 2015).

References:

1. Aziz, RK, D Bartels, AA Best, M DeJongh, T Disz, RA Edwards, ... and O Zagnitko. 2008. The RAST Server: rapid annotations using subsystems technology. BMC genomics 9(1) pp75.
2. Balashov, Sergei P., et al. "Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna." *Science* 309.5743 (2005): 2061-2064.
3. Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., & Madden T.L. (2008) "BLAST+: architecture and applications." BMC Bioinformatics 10:421.
4. Cavaletti, L., Monciardini, P., Bamonte, R., Schumann, P., Rohde, M., Sosio, M. & Donadio, S. (2006). New lineage of filamentous, spore- forming, gram-positive bacteria from soil. Appl Environ Microbiol 72, 4360–4369.
5. Chang, Yun-Juan, et al. "Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium Ktedonobacter racemifer type strain (SOSP1-21 T)." *Standards in genomic sciences* 5.1 (2011): 97.

6. Cheng, Q., Day, A., Dowson-Day, M., Shen, G. F. & Dixon, R. The Klebsiella pneumoniae nitrogenase Fe protein gene (nifH) functionally substitutes for the chlL gene in Chlamydomonas reinhardtii. *Biochem. Biophys. Res. Commun.* **329,** 966–975 (2005).

7. Dodsworth, J. A. *et al*. Thermoflexus hugenholtzii gen. nov., sp. nov., a thermophilic, microaerophilic, filamentous bacterium representing a novel class in the Chloroflexi, Thermoflexia classis nov., and description of Thermoflexaceae fam. nov. and Thermoflexales ord. nov. *Int. J. Syst. Evol. Microbiol.* **64,** 2119–2127 (2014).

8. Doolittle, R.F. 1986. Of Urfs and Orfs: Primer on how to analyze derived amino acid sequences, University Science Books, Mill Valley, CA, USA.

9. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput
Nucleic Acids Res. 32(5):1792-1797

10. Garrity, G. M. & Holt, J. G. (2001). Phylum BVI. Chloroflexi phy. nov. In Bergey's Manual of Systematic Bacteriology, 2nd edn, vol. 1, pp. 427–446. Edited by D. R. Boone, R. W. Castenholz & G. Garrity. New York: Springer.

11. Gouy, M. Guindon, S. & Gascuel., O. (2010) SeaView version 4 : a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution 27(2):221-224.

12. Gupta, RS, P Chander, and S George. 2013. Phylogenetic framework and molecular signatures for the class Chloroflexi and its different clades;

proposal for division of the class Chloroflexi class. Nov. into the suborder Chloroflexineae subord. nov., consisting of the emended family Oschillochloridaceae and the family Chloroflexaceae fam. nov., and the suborder Roseiflexineae subord. nov., containing the family Roseiflexaceae fam nov. Antonie van Leeuwenhoek 103 pp99-119.

13. Hanada, S, A Hiraishi, K Shimada, and K Matsuura. 1995. Chloroflexus aggregans sp. nov., a filamentous phototrophic bacterium which forms dense cell aggregates by active gliding movement. IJSEM 45(4) pp676-681.

14. Hanada, Satoshi. "The Phylum Chloroflexi, the Family Chloroflexaceae, and the Related Phototrophic Families Oscillochloridaceae and Roseiflexaceae." *The Prokaryotes*. Springer Berlin Heidelberg, 2014. 515-532.

15. Hansmann, Sabine, and William Martin. "Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis." *International Journal of Systematic and Evolutionary Microbiology* 50.4 (2000): 1655-1663.

16. Hemp J, Ward LM, Pace LA, Fischer WW. 2015a. Draft genome sequence of Ornatilinea apprima P3M-1, an anaerobic member of the Chloroflexi class Anaerolineae. Genome Announc 3(6):e01353-15.

17. Hemp J, Ward LM, Pace LA, Fischer WW. 2015b. Draft genome sequence of Ardenticatena maritima 110S, a thermophilic nitrate- and iron-reducing

member of the Chloroflexi class Ardenticatenia. Genome Announc 3(6):e01347-15.

18. Hemp J, Ward LM, Pace LA, Fischer WW. 2015c. Draft genome sequence of Levilinea saccharolytica KIBI-1, a member of the Chloroflexi class Anaerolineae. Genome Announc 3(6):e01357-15.

19. Hug, LA, CJ Castelle, KC Wrighton, BC Thomas, I Sharon, KR Frischkorn, KH Williams, SG Tringe, and JF Banfield. 2013. Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. Microbiome 1 pp1-22.

20. Hug, L. A. et al. A new view of the tree and life's diversity. Nat. Microbiol. (2016). doi:10.1038/nmicrobiol.2016.48

21. Hugenholtz, P. & Stackebrandt, E. (2004). Reclassification of Sphaerobacter thermophilus from the subclass Sphaerobacteridae in the phylum Actinobacteria to the class Thermomicrobia (emended description) in the phylum Chloroflexi (emended description). Int J Syst Evol Microbiol 54, 2049–2051.

22. Igarashi N, Harada J, Nagashima S, Matsuura K, Shimada K, Nagashima KVP. 2001. Horizontal Transfer of the Photosynthesis Gene Cluster and Operon Rearrangement in Purple Bacteria. J. Mol. Evol. 52(4):333–41

23. Inagaki, F, M Suzuki, K Takai, H Oida, T Sakamoto, K Aoki, KH Nealson, and K Horikoshi. 2003. Microbial communities associated with geological

horizons in coastal subseafloor sediments from the Sea of Okhotsk. Appl. Environ. Microbiol. 69(12) pp7224-7235.

24. Kaneoka, Ichiro, Naoki Isshiki, and Shigeo Zashu. "K-Ar ages of the Izu-Bonin islands." *Geochemical Journal* 4.2 (1970): 53-60.

25. Kawaichi, S., Ito, N., Kamikawa, R., Sugawara, T., Yoshida, T. & Sako, Y. (2013). Ardenticatena maritima gen. nov., sp. nov., a ferric iron- and nitrate-reducing bacterium of the phylum 'Chloroflexi' isolated from an iron-rich coastal hydrothermal field, and description of Ardenticatenia classis nov. Int J Syst Evol Microbiol 63, 2992–3002.

26. Kawaichi, Satoshi, et al. "Draft genome sequence of a heterotrophic facultative anaerobic thermophilic bacterium, Ardenticatena maritima strain 110ST." *Genome announcements* 3.5 (2015): e01145-15.

27. Klatt, C. G. et al. Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J. 5, 1262–1278 (2011).

28. Klueglein, N., and A. Kappler. "Abiotic oxidation of Fe (II) by reactive nitrogen species in cultures of the nitrate-reducing Fe (II) oxidizer Acidovorax sp. BoFeN1–questioning the existence of enzymatic Fe (II) oxidation."Geobiology 11.2 (2013): 180-190.

29. Kubo, K, K Knittel, R Amann, M Fukui, and K Matsuura. 2011. Sulfur-metabolizing bacterial populations in microbial mats of the Nakabusa hot spring, Japan. Systematic and Applied Microbiology 34(4) pp293-302.

30. Letunic, Ivica, and Peer Bork. "Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees." Nucleic acids research 44.W1 (2016): W242-W245.

31. Li, D., Luo, R., Löffler, C.M., Leung, C.M., Ting, H.F., Sadakane, K., Yamashita, H. and Lam, T.W., 2016. MEGAHIT v1.0: A Fast and Scalable Metagenome Assembler driven by Advanced Methodologies and Community Practices. Methods.

32. Löffler, F. E., Yan, J., Ritalahti, K. M., Adrian, L., Edwards, E. A., Konstantinidis, K. T., Mu¨ ller, J. A., Fullerton, H., Zinder, S. H. & Spormann, A. M. (2013). Dehalococcoides mccartyi gen. nov., sp. nov., obligately organohalide-respiring anaerobic bacteria relevant to halogen cycling and bioremediation, belong to a novel bacterial class, Dehalococcoidia classis nov., order Dehalococcoidales ord. nov. and family Dehalococcoidaceae fam. nov., within the phylum Chloroflexi. Int J Syst Evol Microbiol 63, 625–635.

33. Majumder, Erica LW, Jeremy D. King, and Robert E. Blankenship. "Alternative Complex III from phototrophic bacteria and its electron acceptor auracyanin." *Biochimica et Biophysica Acta (BBA)-Bioenergetics* 1827.11 (2013): 1383-1391.

34. McInerney, James O., Mary J. O'connell, and Davide Pisani. "The hybrid nature of the Eukaryota and a consilient view of life on Earth." *Nature Reviews Microbiology* 12.6 (2014): 449-455.

35. Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA pp 1 - 8.

36. Moe, W. M., Yan, J., Nobre, M. F., da Costa, M. S. & Rainey, F. A. (2009). Dehalogenimonas lykanthroporepellens gen. nov., sp. nov., a reductively dehalogenating bacterium isolated from chlorinated solvent-contaminated groundwater. Int J Syst Evol Microbiol 59, 2692–2697.

37. Mulkidjanian, A. Y. *et al*. The cyanobacterial genome core and the origin of photosynthesis. *Proc Natl Acad Sci U S A* **103,** 13126–13131 (2006).

38. Nagashima S, Nagashima KVP. 2013. Comparison of Photosynthesis Gene Clusters Retrieved from Total Genome Sequences of Purple Bacteria. In Genome Evolution of Photosynthetic Bacteria, pp. 151–78. Academic Press

39. Overmann, J. 2008. Green nonsulfur bacteria. In: Encyclopedia of Life Sciences (ELS). John Wiley & Sons, Ltd: Chichester.

40. Oyaizu, H., et al. "The green non-sulfur bacteria: a deep branching in the eubacterial line of descent." *Systematic and applied microbiology* 9.1-2 (1987): 47-53.

41. Pace, Norman R. "Mapping the tree of life: progress and prospects." *Microbiology and Molecular Biology Reviews* 73.4 (2009): 565-576.

42. Pace LA, Hemp J, Ward LM, Fischer WW. 2015. Draft genome of Thermanaerothrix daxensis GNS-1, a thermophilic facultative anaerobe from the Chloroflexi class Anaerolineae. Genome Announc 3(6):e01354-15.

43. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2014. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research, 25: 1043-1055.

44. Pierson, B. K. & Castenholz, R. W. A Phototrophic Gliding Filamentous Bacterium of Hot Springs , Chloroflexus aurantiacus , gen . and sp. nov. Arch. Microbiol. 100, 5–24 (1974).

45. Raymond, J., Zhaxybayeva, O., Gogarten, J. P., Gerdes, S. Y. & Blankenship, R. E. Whole-Genome Analysis of Photosynthetic Prokaryotes. Science (80-.). 298, (2002).

46. Seviour RJ, Blackall LL. 1999. Current taxonomic status of filamentous bacteria found in activated sludge plants, pp. 122–146. In The Microbiology of Activated Sludge. Springer Netherlands, Dordrecht.

47. Sharma, A. K., Zhaxybayeva, O., Papke, R. T. & Doolittle, W. F. Actinorhodopsins: Proteorhodopsin-like gene sequences found predominantly in non-marine environments. *Environ. Microbiol.* **10,** 1039–1056 (2008).

48. Sorokin, D. Y. *et al*. Nitrification expanded: discovery, physiology and genomics of a nitrite-oxidizing bacterium from the phylum Chloroflexi. *ISME J*. **6,** 2245–2256 (2012).

49. A. Stamatakis: "RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies". In Bioinformatics, 2014

50. Stookey, Lawrence L. "Ferrozine---a new spectrophotometric reagent for iron." Analytical chemistry 42.7 (1970): 779-781.

51. Strous M, Kraft B, Bisdorf R, TegetMeyer H (2012) The binning of metagenomic contigs for microbial physiology of mixed cultures. Frontiers in Microbial Physiology and Metabolism 3:410.

52. Ward LM, Hemp J, Pace LA, Fischer WW. 2015a. Draft genome sequence of Leptolinea tardivitalis YMTK-2, a mesophilic anaerobe from the Chloroflexi class Anaerolineae. Genome Announc 3(6):e01356-15.

53. Ward LM, Hemp J, Pace LA, Fischer WW. 2015b. Draft genome sequence of Herpetosiphon geysericola GC-42, a nonphototrophic member of the Chloroflexi class Chloroflexia. Genome Announc 3(6):e01352-15

54. Ward, LM, A Idai, WW Fischer, and SE McGlynn. 2017a. Microbial diversity and productivity of Jinata Onsen, an iron-rich intertidal hot spring in Japan.

55. Ward, LM, WW Fischer, K Matsuura, and SE McGlynn. 2017b. Cone-forming microbial mats of Nakabusa Onsen, Japan as analogs of Precambrian stromatolites. In prep.

56. Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench" Bioinformatics25 (9) 1189-1191 doi: 10.1093/bioinformatics/btp033

57. Wätzlich, D. *et al*. Chimeric Nitrogenase-like enzymes of (bacterio)chlorophyll biosynthesis. *J. Biol. Chem*. **284,** 15530–15540 (2009).

58. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).

59. Woese CR, Debrunnervossbrinck BA, Oyaizu H, Stackebrandt E, Ludwig W (1985) Gram-positive bacteria—possible photosyn- thetic ancestry. Science 229(4715):762–765

60. Woese, Carl R. "Bacterial evolution." *Microbiological reviews* 51.2 (1987): 221.

61. Xiong, J, WM Fischer, K Inoue, M Nakahara, and CE Bauer. 2000. Molecular evidence for the early evolution of photosynthesis. Science 289 pp1724-1730.

62. Yabe, S., Aiba, Y., Sakai, Y., Hazaka, M. & Yokota, A. (2010). Thermosporothrix hazakensis gen. nov., sp. nov., isolated from compost, description of Thermosporotrichaceae fam. nov. within the class Ktedonobacteria Cavaletti et al. 2007 and emended description of the class Ktedonobacteria. Int J Syst Evol Microbiol 60, 1794– 1801.

63. Yamada M, Zhang H, Hanada S, Nagashima KVP, Shimada K, Matsuura K. (2005). Structural and spectroscopic properties of a reaction center complex from the chlorosome-lacking filamentous anoxygenic phototrophic bacterium *Roseiflexus castenholzii*. *J Bacteriol* **187**: 1702–1709.

64. Yamada, T., Sekiguchi, Y., Hanada, S., Imachi, H., Ohashi, A., Harada, H. &

Kamagata, Y. (2006). Anaerolinea thermolimosa sp. nov., Levilinea saccharolytica gen. nov., sp. nov. and Leptolinea tardivitalis gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes Anaerolineae classis nov. and Caldilineae classis nov. in the bacterial phylum Chloroflexi. Int J Syst Evol Microbiol 56, 1331–1340.

65. Yamada, T and Y Sekiguchi. 2009. Cultivation of uncultured Chloroflexi subphyla: significance and ecophysiology of formerly uncultured Chloroflexi 'Subphylum I' with natural and biotechnological relevance. Microbes Environ. 24pp205–216.

66. Youvan DC, Bylina EJ, Alberti M, Begusch H, Hearst JE. (1984). Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from *R. capsulata*. *Cell* **37**: 949–957.

67. Yanyushin MF, del Rosario MC, Brune DC, Blankenship RE. 2005. New class of bacterial membrane oxidoreductases. Biochemistry (Mosc.). 44(30):10037–45 Zahnle

68. Zeng, Y., Feng, F., Medova, H., Dean, J. & Kobli ek, M. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum Gemmatimonadetes. Proc. Natl. Acad. Sci. 111, 7795–7800 (2014)

Supplementary information:

Kouleothrix aurantiaca

Kouleothrix aurantiaca, a member of the group formerly known as 'Eikenboom morphotype 1851' (Seviour and Burkall 1999), was isolated from activated sludge in an industrial wastewater treatment facility (Kohno et al. 2002). It forms orange-pigmented cells organized into long mm-scale filaments, grows on numerous sugars and pyruvate and by fermentation on certain sugars, and can reduce nitrate to nitrite (Kohno et al. 2002). It is closely related to members of the genus *Roseiflexus* (Beer et al. 2002), however phototrophy has not been observed in these organisms.

The genome of *Kouleothrix aurantiaca* COM-B (JCM 19913) was sequenced as part of a project to expand the phylogenetic breadth of *Chloroflexi* genomes. Genome sequencing was performed at Seqmatic using the Illumina MiSeq sequencing platform. SPAdes 3.1.1 (7) was used to assemble the genome. The genome was screened for contaminants based on sequence coverage, GC composition, and BLAST hits of conserved single copy genes. Genome annotation was performed using the NCBI Prokaryotic Genome Annotation Pipeline.

K. aurantiaca falls within the phototrophic Chloroflexia, with a consistent position basal to *Roseiflexus* in both organismal and photosynthetic gene trees (Figures 2 and 3). *K. aurantiaca* encodes for all of the genes required for anoxygenic phototrophy; a type II reaction center (including a fused pufLM and a pufC), a complete bacteriochlorophyll biosynthesis pathway, and a cytochrome bc complex, but no Alternative Complex III. *K. aurantiaca* also encodes for a branched aerobic respiration pathway, including two

Complex Is (NADH dehydrogenase), Complex II (succinate dehydrogenase), two Complex IIIs (cytochrome bc complex), two heme-copper oxygen reductases (A and B-family), and a quinol bd oxidase. In addition it has a NirK nitrite reductase. *K. aurantiaca* encodes a type 1 rubisco and a phosphoribulokinase gene, suggesting that it is capable of carbon fixation via the Calvin Cycle. It does not, however, encode key genes in the 3-hydroxypropionate bicycle used for carbon fixation in *Chloroflexus* and *Roseiflexus* (Klatt et al. 2007).

Supplemental discussion

CP2_42A also encodes a rhodopsin homolog most closely related to xanthorhodopsin, a proton-pumping rhodopsin shown to use light-harvesting antenna carotenoids (Balashov et al. 2005). CP2_42A also encodes for a branched aerobic respiration pathway, including two Complex Is (NADH dehydrogenase), Complex II (succinate dehydrogenase), Complex III (cytochrome bc complex), Alternative Complex III, and two heme-copper oxygen reductases (A and B-family). The CP2_42A genome bin did not recover genes for the Calvin cycle or the 3-hydroxypropionate bicycle, suggesting that it may grow primarily as a photoheterotroph.

JP3_7 also encodes for a branched aerobic respiration pathway, including two Complex Is (NADH dehydrogenase), Complex II (succinate dehydrogenase), Complex III (cytochrome bc complex), two heme-copper oxygen reductases (A and B-family), and a cytochrome bd oxidase.

The bacteriochlorophyll synthesis genes present in JP3_7 and CP2_42A may record a hybrid history of the pathway. While most genes in the pathway (bchX,Y,Z,F,G,P) recovered a close relationship between all of the phototrophic Chloroflexi, the bchI,D, and H (and H-like) genes show a closer relationship between the Chloroflexia and Clorobi to the exclusion of CP2_42A and JP3_7. This may record a hybrid history for bacteriochlorophyll synthesis in the novel Chloroflexi phototrophs, with some genes in the bacteriochlorophyll synthesis pathway being transferred with the reaction center genes from the Chloroflexia, and others deriving from the Chlorobi or other phototrophic bacteria. Alternatively, this may reflect the relative lack of conserved sequence information of short, relatively quickly-evolving proteins relative to the much more conserved and information-rich reaction center and ribosomal proteins.

Phylogenies of electron transport proteins reveal that aerobic respiration using an A family heme copper oxidoreductase and a *bc* complex is a vertically-inherited synapomorphy of the Thermofonseae, while the B family heme copper oxidoreductase and Alternative Complex III found in CP2_42A appear to have been acquired through horizontal gene transfer associated with the acquisition of the type 2 reaction center.

Genes involved in lipopolysaccharide synthesis (e.g. lpxB, lpxC, kdsA) and outer membrane proteins (e.g. bamA) were absent from all Chloroflexi genomes reported here. This is consistent with the proposed single membrane "monoderm" nature of Chloroflexi (Sutcliffe 2010, Sutcliffe 2011) and suggests that this is a conserved feature of the Chloroflexi phylum, though the presence of outer membrane proteins and lipopolysaccharide synthesis in the closely related Armatimonadetes phylum (e.g. Ward et

al. 2017c) suggests that monoderm Chloroflexi are derived from a diderm ancestor and are not representative of an ancestral state.

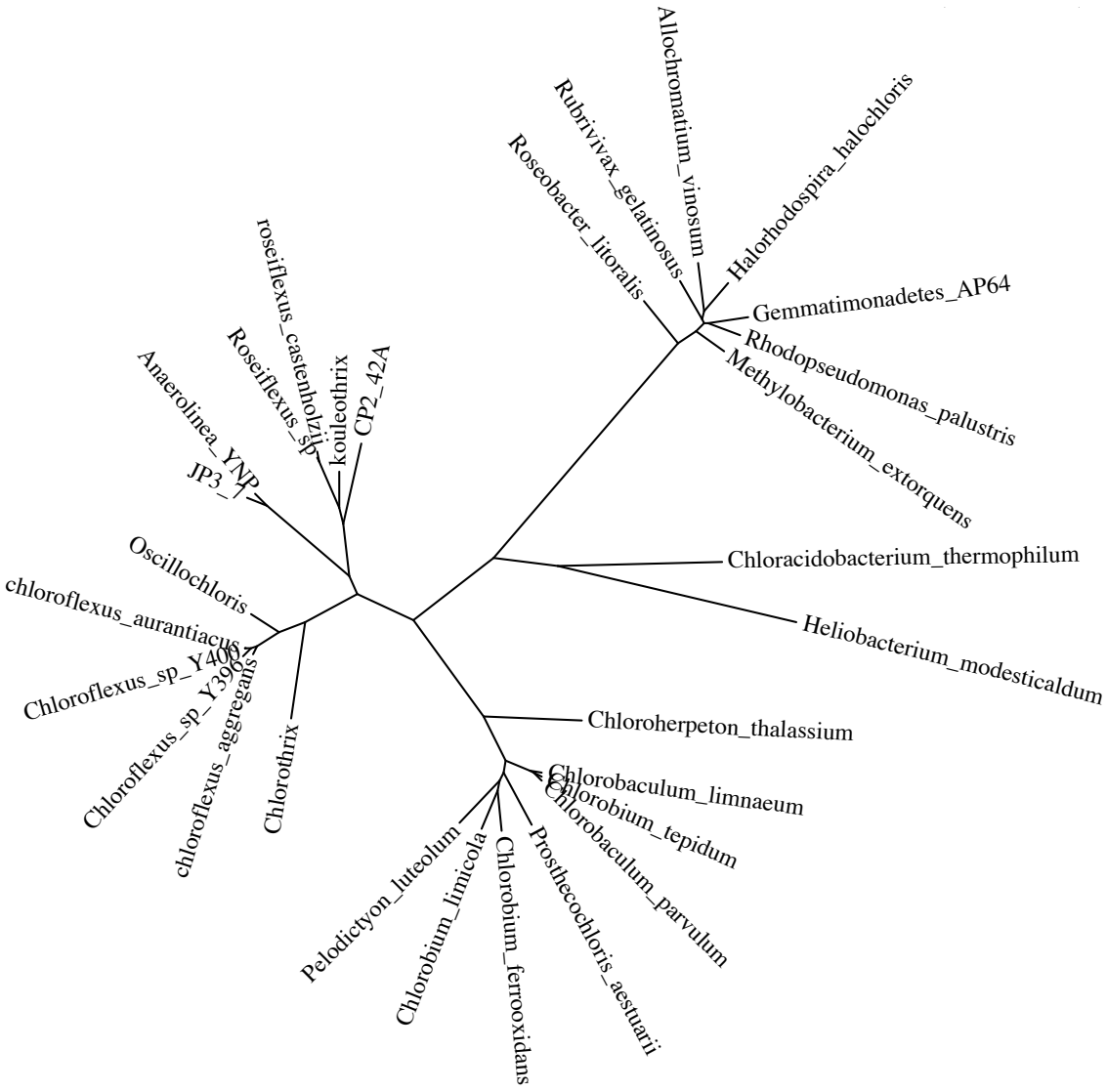
Probability of missing genes

In order to estimate the probability that genes were missing from recovered genome bins, we calculated the probability mass function of recovering zero genes of a particular set from a genome of predicted size, given estimated completeness and assuming random sampling without replacement of individual genes. Though gene size varies significantly and colocalization makes selection of related genes not entirely independent, we assume here that all genes have an equal probability of being selected. This simplifying assumption is reasonable, as recovered phototrophy genes largely reside on separate contigs (suggesting that colocalization is limited) and the length of phototrophy-related genes (e.g. coding for reaction center proteins, bacteriochlorophyll synthases, etc) are within error of average gene length. The calculation took the form of $f(x) = \binom{n}{x}\binom{T-n}{r-x}/\binom{T}{r}$, where f is the probability of recovering x genes of set r from a genome containing T genes of which n are recovered. In the case of our genome bins, n equaled the number of protein coding sequences recovered in each bin, T equaled n divided by the completeness of the genome as estimated by CheckM, and r equaled 6 (representing pufL, pufM, pufC, bchX, bchY, and bchZ). The probability that phototrophy genes existed in in *Ca*. Thermofonseae genomes but was not recovered in our bins ranged from ~0.5 for JP1_191 (at only ~10% completeness) to ~2*10^-13 for JP3_13 (at over 96% completeness). The probability of missing phototrophy genes was only >5% in JP1_191, greatly improving our confidence

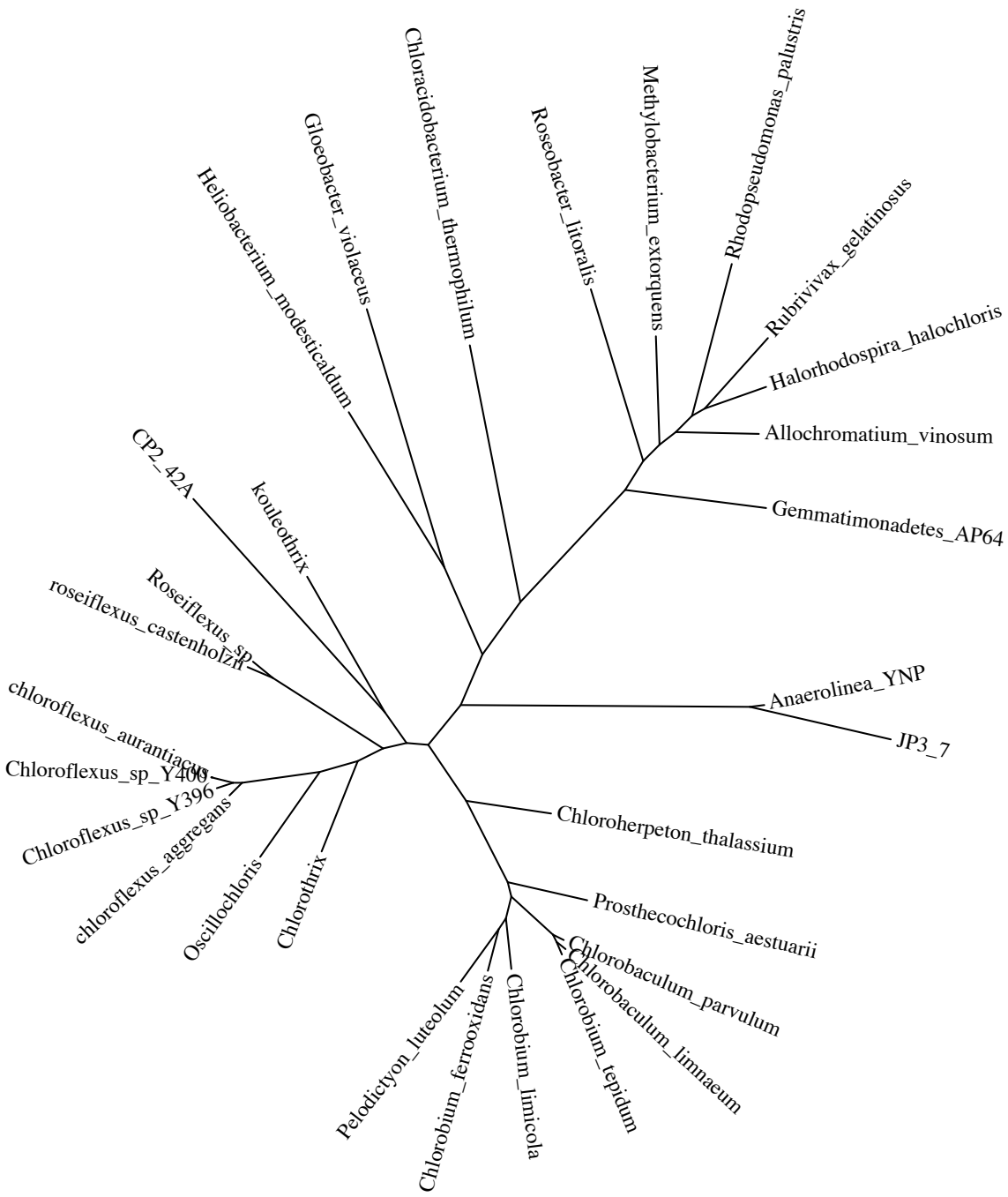
that the absence of phototrophy from most strains of *Ca*. Thermofonseae is a real signal and not an artifact of incomplete genome bins.

A similar calculation can be made for the probability that bchL, N, B, M, or E genes are present in phototrophic Thermofonseae but simply not recovered in our genome bins. The probability of missing all five of these genes is about 0.03% for CP2_42A and less than 0.005% for JP3_7, increasing the possibility that these genes are in fact missing from these genomes, potentially replaced by promiscuous homologs.

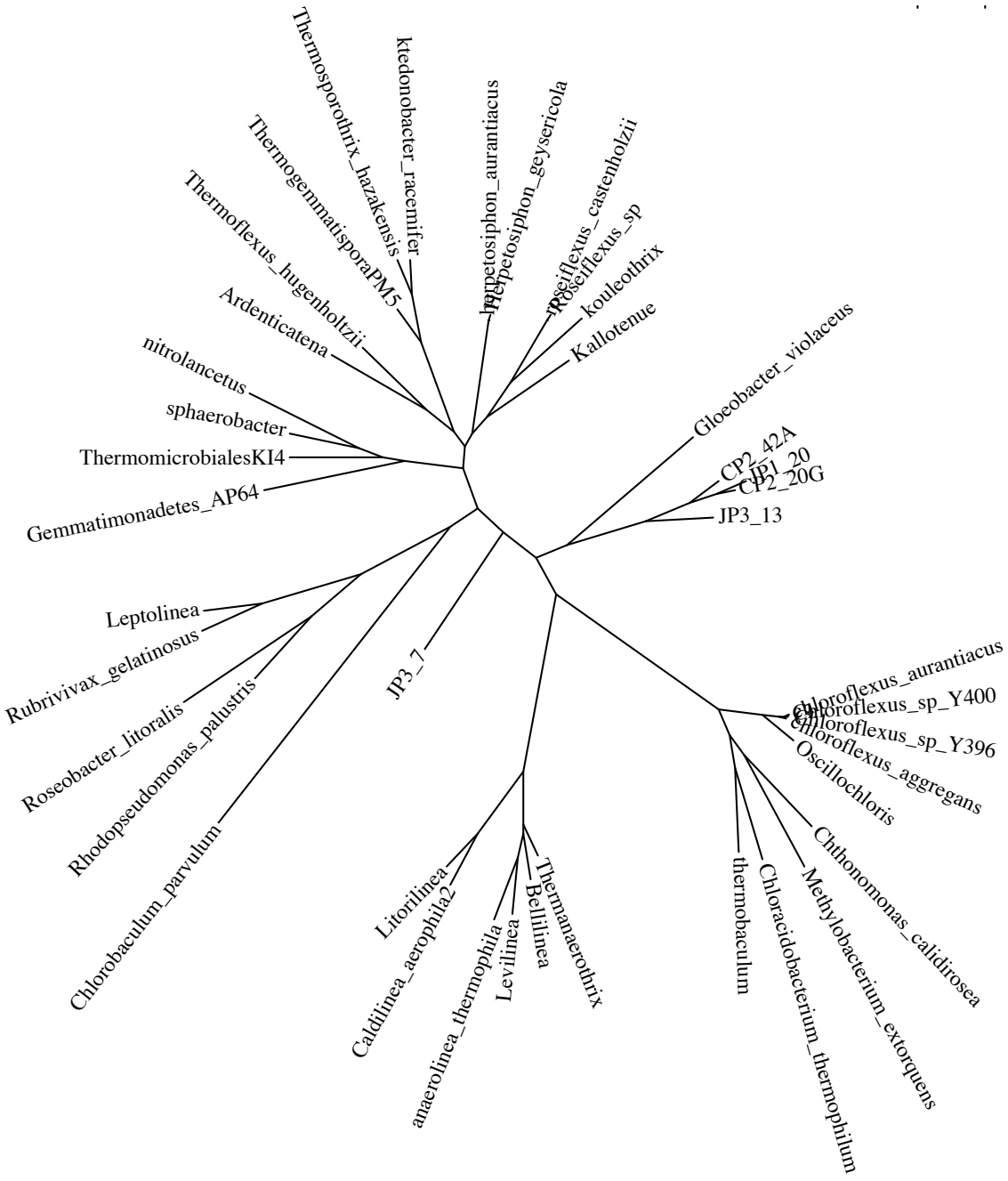
Supplemental figure 1: Unrooted phylogeny of bchXYZ protein sequences



Supplemental figure 2: Unrooted phylogeny of bchIDH protein sequences

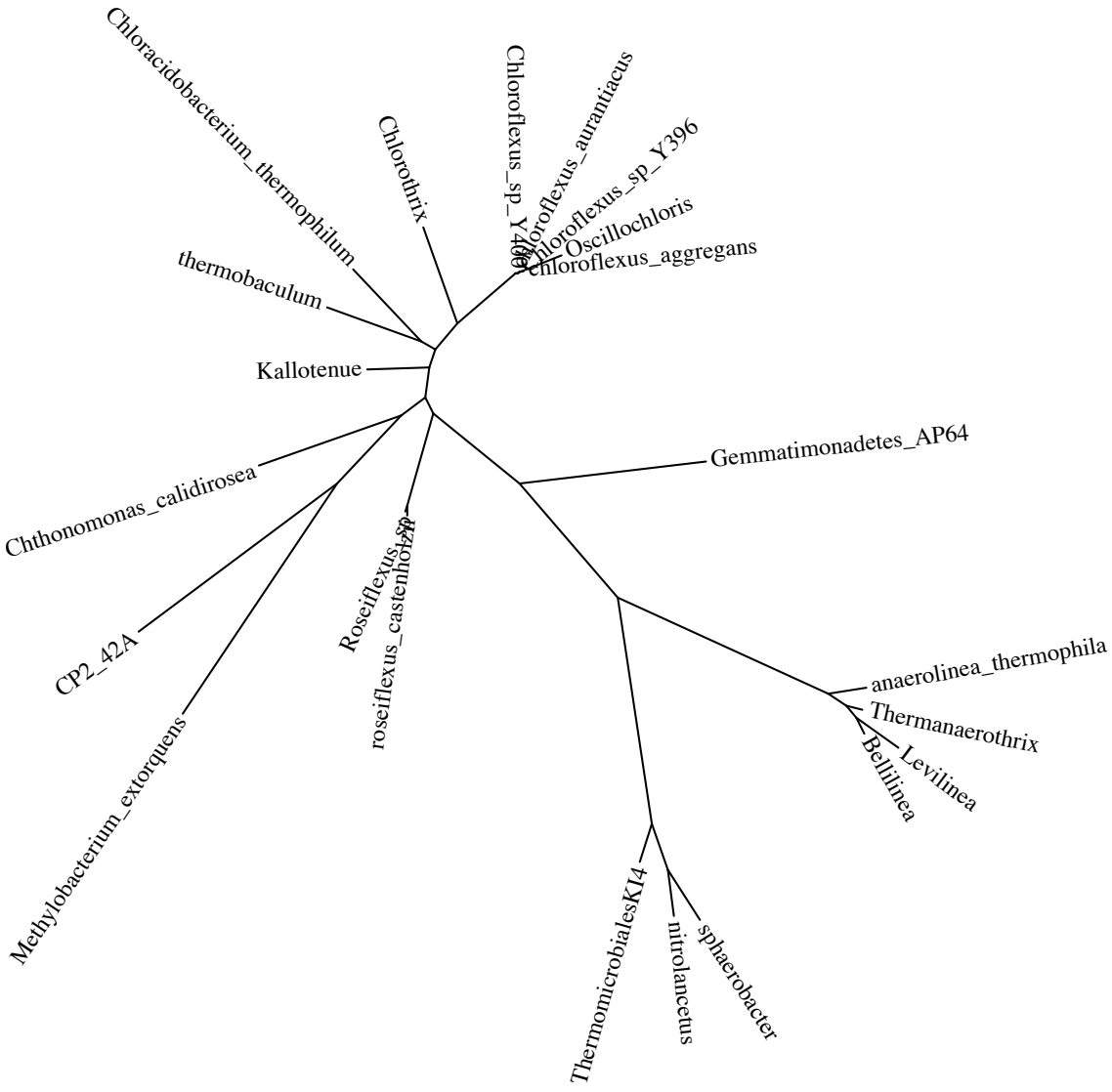


Supplemental figure 3: Unrooted phylogeny of A-family Heme Copper Oxidoreductase protein sequences

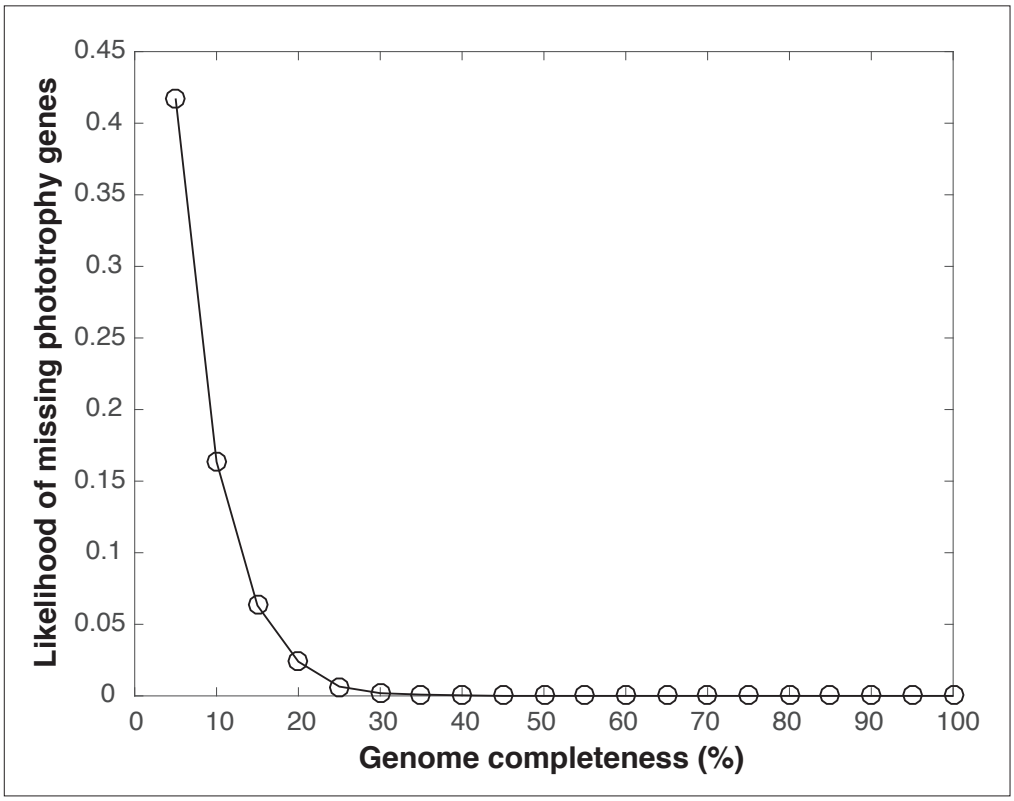


Supplemental figure 4: Unrooted phylogeny of B-family Heme Copper Oxidoreductase protein sequences

Supplemental figure 5: Unrooted phylogeny of Alternative Complex III protein sequences



Supplemental figure 6: Unrooted phylogeny of bc complex protein sequences



Supplemental figure 7: Probability of failure to recover phototrophy genes for a given completeness of genome recovery. Results plotted here are for a simulation following the constraints and logic discussed in the text.

Supplemental Table 1: Genome statistics of sequenced strains

	Genome size	% GC	# coding sequences	# Contigs	Completeness	Contamination	tRNAs	Source
CP1_1M	1.39	59	1182	138	42.28	1.81	14	Nakabusa Cone Pool 1
CP2_2F	1.99	59	1734	20	49.46	0	23	Nakabusa Cone Pool 2
CP2_20G	3.09	48	2678	852	78.54	3.55	32	Nakabusa Cone Pool 2
CP2_42A	3.3	59	2897	2024	79.44	10.42	31	Nakabusa Cone Pool 2
JP1_8	2.21	51	1973	601	58.13	0.13	17	Jinata Pool 1
JP1_16	4.06	44	3238	1764	95.15	17.31	45	Jinata Pool 1
JP1_20	3.36	46	2878	1139	79.09	4.78	34	Jinata Pool 1
JP1_191	0.417	47	334	883	10.63	1.8	7	Jinata Pool 1
JP3_7	3.62	63	3078	1331	87	12.85	46	Jinata Pool 3
JP3_13	3.67	60	3116	1259	96.17	10.87	46	Jinata Pool 3
Kouleothrix aurantiaca	8.7	62	8993	5539	85	0	97	Isolate from wastewater sludge

Supplemental Table 2: Metabolic traits of genomes reported here

	16S	RpoB	RCII	bc	ACII	Rhodopsin	Denitrification	A Fam HCO	B Fam HCO	3HP	Calvin Cycle
CP1_1M	-	+	-	+	-	-	-	-	-	-	-
CP2_2F	-	+	-	-	-	+	-	-	-	-	-
CP2_20G	-	+	-	+	-	-	nirK	+	-	-	-
CP2_42A	-	+	+ (fused)	+	+	+	-	+ (two)	+	-	-
JP1_8	-	+	-	-	-	-	-	-	-	-	-
JP1_16	-	+	-	+	-	-	-	+	-	-	-
JP1_20	-	+	-	+	-	-	nirK, NOR	+ (three)	-	-	-
JP1_191	-	+	-	-	-	-	-	-	-	-	-
JP3_7	-	+	+ (unfused)	+	-	-	-	+	+	-	-
JP3_13	-	+	-	+	-	+	-	+ (two)	-	-	-
Kouleothrix aurantiaca	+	+	+ (fused)	+	-	-	nirK	+	+	-	+

Supplemental references:

1. Beer M, Seviour EM, Kong Y, Cunningham M, Blackall LL, Seviour RJ. 2002. Phylogeny of the filamentous bacterium Eikelboom Type 1851, and design and application of a 16S rRNA targeted oligonucleotide probe for its fluorescence in situ identification in activated sludge. FEMS Microbiol Lett 207:179–183.
2. Klatt, Christian G., Donald A. Bryant, and David M. Ward. "Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats." *Environmental microbiology* 9.8 (2007): 2067-2078.
3. Kohno T, Sei K, Mori K. 2002. Characterization of type 1851 organism isolated from activated sludge samples. Water Sci Technol 46:111–114.
4. Sutcliffe, I. C. A phylum level perspective on bacterial cell envelope architecture. *Trends Microbiol*. **18,** 464–470 (2010).
5. Sutcliffe, I. C. Cell envelope architecture in the Chloroflexi: A shifting frontline in a phylogenetic turf war. *Environmental Microbiology* **13,** 279–282 (2011).
6. Ward, LM, SE McGlynn, and WW Fischer. 2017c. Draft genomes of two members of the Armatimonadetes phylum recovered from Japanese hot spring metagenomes. In prep.